typesetting breakthrough

HYPHENLESS JUSTIFICATION

by GEORGE E. KUNKEL and TILMON H. MARCUM

While an increasingly large number of computer and printing organizations continue to struggle with the hyphenation problem in computer produced text, a small team of specialists in the Central Intelligence Agency has developed a novel approach which is amazing in its simplicity and which may make the troublesome end-of-line hyphen a thing of the past in computerized photocomposition.

The success or failure of present computerized book composition systems hinges largely upon the computer's ability to provide correct end-of-line word divisions required for justification. Several years and hundreds of thousands of dollars have been spent in attempting to develop dictionaries and logic which will attain the necessary accuracy required for high quality composition. Existing computer systems cannot yet equal the accuracy of the average keyboard operator in hyphenating end-of-line words although a fairly high degree of accuracy is obtained in some cases. Accuracy is needed in computer composition because of the problem of correcting errors and the constant demand for greater speeds.

With the present state of the art, the printer must choose between a system obtaining high accuracy with proportionately higher computer costs or a less sophisticated system providing a lesser degree of accuracy with the attendant error cost factor. Any system permitting errors poses the problem of new errors being introduced when corrections are required.

The idea of eliminating the end-of-line hyphen is not new. Some newspapers and printers now justify text without hyphens by the use of excessive interword spacing and fixed letter spacing or a combination of both. Some publications appear with an unjustified single or multi-column format. Unfortunately neither of these systems provides acceptable typographic quality for book production.

A primary rule in typesetting is to avoid hyphens wherever possible since they destroy continuity in reading. It is obvious, then, that the real need is for a simple plan which eliminates the end-of-line hyphen without sacrificing typographic quality. Uncle Sam's team feels that they have solved this problem by the use of a variable set size technique on a line-for-line basis.* In simpler terms, this means that a sort of "coefficient of expansion" or contraction is applied to the proportional spacing between characters in each line of text. The line is thus expanded and contracted without destroying the proportional values of the individual character as is the case with fixed letter-spacing. The variation in appearance of the lines of text produced by this method is sufficiently subtle to remain unnoticed by the average reader. The ability to expand and contract provides sufficient latitude in justification so that the need for end-of-line word division is rare. The computer programming and processing which is required for variable set size justification is many times simpler and less costly than that required for end-of-line word division and hyphenation.

In a unit font each character is assigned a proportional unit value, which is valid regardless of the set size. The specified line measure can then be stated in units for each set size. Therefore the computer can accumulate units as it passes through four overlapping (set sizes) zones of justification in which an interword space may be selected for the end of line. This simple logic for line justification thus eliminates the requirements for stored dictionaries for word hyphenation, programmed logic approaches, or a combination of the two, prefix and suffix tables, and stripping and reconstituting routines.

The Central Intelligence Agency prints in its own facilities a number of high quality book-type publications. These publications are set in type on Intertype and Monotype machines and are printed by offset from plates made from positive Mylar proofs pulled directly from the type. For the past five years the majority of this composition has been done on Model F4 Intertype machines operated by TTS tape.

About a year ago, the agency printing engineers and computer specialists began studying the possibility of doing page composition by photocomposing from computer-prepared tape. This study developed the feasibility of this type of composition and a thorough investigation was made of all available photocomposing equipment. After a great deal of study a Model 513 Photon has been obtained for this composition. A primary reason for this selection was the 513's capability of changing set sizes from codes inserted into the computer-prepared tape.

The printing engineers and computer specialists began the job of programming an IBM 1410 computer for this typesetting job. Agency personnel had talked with other

*Set size determines the overall horizontal dimension of the space assigned to a character—it includes space for separation from adjacent letters. Set size also expresses the relationship of one type font to another. The size relationship of one letter to another is expressed in "units of relative value," based on the size of the em (18/18). Therefore the horizontal dimension for a letter is determined:

$$URV \times SS = \text{Horizontal dimension}$$

$$\text{Point Size} \left(\frac{1''}{72} \right) = \text{Vertical dimension}$$

Continued

printers and computer personnel and had visited numerous printing plants in their investigation of the problems involved in this method of typesetting. Since they were interested in page composition this aspect of the problem received most of the initial attention. It soon became apparent, however, that the primary problem involved in this type of computerized composition was not in page makeup but rather end-of-line word division and hyphenation. Consequently the computer specialists and printing engineers gave consideration to the number of known methods of handling this problem. It was apparent that the end-of-line hyphenating problem would require considerable programming time and computer capacity. Further, it was found that hyphenating inaccuracies, computer capacity, loss of speed, cost of dictionaries, etc., were problems not yet overcome.

Before getting into programming for hyphenation, the agency personnel began a thorough study of any alternatives to end-of-line hyphenation. Since the 513 Photon had been selected for the composing job because of its tape-operated set size changing capability, the agency engineers felt there was a possibility of using this capability for justification without end-of-line hyphens in average book composition. Further study indicated that hyphenless justification of practically any measure of composition was possible by varying set sizes of individual lines.

The varying of set sizes permits a delicate method of proportional letter spacing. Numerous sample pages have been set in which four different set sizes are used and the resulting typography is satisfactory for book composition of relatively high quality. Since the program is not yet operative the described technique was simulated on a Monotype keyboard and the sample pages set on a Monophoto caster.

The system of employing proportionate letter spacing by set size changes will work as follows:

The computer is being programmed to compute for end-of-line decisions utilizing two, three, four or five set sizes close enough in size to each other to be inoffensive typographically and yet disparate enough to significantly increase justification range. The computer will arrive at end-of-line decisions for all set sizes involved, choose the set size in which an interword or other natural line ending code (period, em quad, compounding hyphen, etc.) occurred in the justification range, and incorporate in the output tape the appropriate set size codes to accompany the line.

Many rules and variations in set sizes and data blocks can be developed but the ultimate result is that hyphens can be eliminated or reduced to the degree that they would no longer pose a problem to the computer, the printer, or the reader.

Adoption of this system of typesetting will:

1. Reduce computer processing time.
2. Simplify hyphenation logic and reduce initial programming time and costs.
3. Enable the use of less expensive computer equipment.

This system of phototypesetting from computer-prepared tape has the advantage of completely eliminating the need for the complicated programming now in use for end-of-line hyphenations. This in many instances would indicate that a smaller computer or less peripheral gear may be used for relatively complicated composition. The drastic reduction in the correction problems caused by improper hyphenation is another advantage which many printers will welcome. This is particularly true in the case of photocomposition and its attendant correction problems. Obviously the computer when processing data for this system will need a routine to take care of the inevitable line which will not succumb to the set-size-change method. This routine need only provide for enough expansion and contraction capability in the two lines involved to allow for application of a human decision. A simple solution would be for the computer to produce the lines involved by inserting a hyphen at any point in the troublemaking word which would allow the median set size and a median width interword space. Even with this method the division may be correct, and at least there would always be sufficient latitude for adjustment.

The agency plans to edit tape on the 1410 computer by proofreading one or two computer printouts prior to creation of final tape for use on the photocomposing machine. Thus the rare instance where an end-of-line hyphen would be necessary can be determined from the printout

Continued

Fig. 1

Fig. 2

b. Close Support Operations. The lack of concealment, great distances involved, and mobility of forces—each characteristic of desert operations—necessitate increased emphasis on the employment of tactical air in close support of ground operations. The lack of natural cover and concealment makes for ease of target location and provides better than normal conditions for high-level bombing. Installations stand out due to the contrast between regularly shaped objects and the open barrenness of the desert. Movement is readily apparent from the air because of the dust created and the prominence of shadows. Low-level attacks are handicapped by lack of covered approaches; however, this is offset by the increased visibility which enables aircraft to initiate their firing runs from a greater distance. This improved visibility, coupled with the rapid movement, lack of prominent terrain features, and the fluid situations characteristic of desert operations, necessitates positive action to identify friendly

b. Close Support Operations. The lack of concealment, great distances involved, and mobility of forces—each characteristic of desert operations—necessitate increased emphasis on the employment of tactical air in close support of ground operations. The lack of natural cover and concealment makes for ease of target location and provides better than normal conditions for high-level bombing. Installations stand out due to the contrast between regularly shaped objects and the open barrenness of the desert. Movement is readily apparent from the air because of the dust created and the prominence of shadows. Lowlevel attacks are handicapped by lack of covered approaches; however, this is offset by the increased visibility which enables aircraft to initiate their firing runs from a greater distance. This improved visibility, coupled with the rapid movement, lack of prominent terrain features, and the fluid situations characteristic of desert operations, necessitates positive action to identify friendly.

JUSTIFICATION . . .

and the necessary correction incorporated in the final tape. It should never be necessary to correct hyphenation after the page has been set.*

The relatively high quality of composition possible by this method is illustrated by the accompanying samples. Fig. 1 shows a column of normal composition, 10-point Modern, 19 picas, one set size. This normal composition in a relatively wide measure (19 picas) requires six end-of-line hyphens. Fig. 2 is the same composition using variable set sizes of 10, 10½, and 11. This method of composing eliminates all end-of-line hyphens—the abnormal lines indicated by bullets, and the set size also indicated. Quality of the composition is good, and the variations in letter spacing are not noticeable to anyone not looking for them. It appears that readability of copy is much improved.

Fig. 3 is an example of the hyphenless technique on an 11-pica measure in 8-point type, using three set sizes—7½, 8, and 8½. Eight set was considered normal. Indicated by bullets in Fig. 4 are those lines which used other than eight as the set size. ■

*The authors are indebted to Robert D. Hicks, Ballard Jamieson, and Robert W. Pearson for guidance and assistance during the evolution of the system described here.

Fig. 3

The American prairies are of two kinds. Those which lie east of the Mississippi are comparatively small, are exceedingly fertile, and are always surrounded by forests. They are susceptible of high cultivation, and are fast becoming settled. They abound in Ohio, Michigan, Illinois, and Indiana. They labor under the disadvantages of a scarcity of wood and water—evils of a serious character, until art has had time to supply the deficiencies of nature. As coal is said to abound in all that region, and wells are generally successful, the enterprise of the emigrants is gradually prevailing against these difficulties.

The second description of these natural meadows lies west of the Mississippi, at a distance of a few hundred miles from that river, and is called the Great Prairies. They resemble the steppes of Tartary more than any other known portion of the world; being, in fact, a vast country, incapable of sustaining a dense population, in the absence of the two great necessaries already named. Rivers abound, it is true; but this region is nearly destitute of brooks and the smaller water courses, which tend so much to comfort and fertility.

The origin and date of the Great American Prairies form one of nature's most majestic mysteries. The general character of the United States, of the Canadas, and of Mexico, is that of luxuriant fertility. It would be difficult to find another portion of the world, of the same extent, which has so little useless land as the inhabited parts of the American Union. Most of the mountains are arable; and even the

Fig. 4

The American prairies are of two kinds. Those which lie east of the Mississippi are comparatively small, are exceedingly fertile, and are always surrounded by forests. They are susceptible of high cultivation, and are fast becoming settled. They abound in Ohio, Michigan, Illinois, and Indiana. They labor under the 8½ disadvantages of a scarcity of wood and water—evils of a serious character, 7½ until art has had time to supply the deficiencies of nature. As coal is said to abound in all that region, and wells are generally successful, the enterprise of the emigrants is gradually prevailing 7½ against these difficulties.

The second description of these natural meadows lies west of the Mississippi, at a distance of a few hundred miles from that river, and is called the Great Prairies. They resemble the steppes of Tartary more than any other known portion of the 8½ world; being, in fact, a vast country, incapable of sustaining a dense population, in the absence of the two 8½ great necessaries already named. Rivers abound, it is true; but this region is nearly destitute of brooks and the smaller water courses, which tend so much to comfort and fertility.

The origin and date of the Great American Prairies form one of nature's most majestic mysteries. The general 7½ character of the United States, of the Canadas, and of Mexico, is that of luxuriant fertility. It would be difficult 7½ to find another portion of the world, of the same extent, which has so little useless land as the inhabited parts of the American Union. Most of the mountains are arable; and even the